Submission in Response to NSF CI 2030 Request for Information

DATE AND TIME: 2017-01-18 09:59:26
REFERENCE NO: 173

Author Names & Affiliations

- Emanuel Gull - University of Michigan, Ann Arbor

Contact Email Address (for NSF use only)

(Hidden)

Research Domain, discipline, and sub-discipline

Condensed matter theory, computational quantum many-body theory

Title of Submission

Need for New Algorithms

Abstract (maximum ~200 words).

Moore's law has led to an exponential growth of computer power. Parallel development, often neglected, enabled exponential speedups due to novel computational algorithms. This speedup is substantially faster than Moore's law, so that today's computers running with thirty years old algorithms will be much slower than thirty year old computers with today's algorithms. This submission outlines the needs for *developers* of next generation algorithms for materials design, which are substantially different from the needs of the next generation application users.

Question 1 Research Challenge(s) (maximum ~1200 words): Describe current or emerging science or engineering research challenge(s), providing context in terms of recent research activities and standing questions in the field.

The quantum many-body problem, i.e. the solution of the Schrodinger equation for a large number of particles and the calculation of quantum observables, governs the physics of all materials. Its exact solution for the wave function of a large system is intractable. The challenge is therefore to devise approximate solutions that are (a) accurate enough to be useful for materials design and experiment, and (b) computationally tractable.

This is not an engineering question where existing equations have to be accelerated, parallelized, or 'made exascale ready'. Rather, it is a theoretical physics question: new equations and approximations need be developed, implemented, tested, and improved. Successful theory develpment in this area has led to the development of practical implementations of the density functional theory (1970s), the GW approximation (1980s), the DMRG and quantum Monte Carlo methods (1990s), the DMFT with continuous time algorithms (2000s), and a recent flurry of activity in many different theory approaches that, for simple models, are now quantitative and accurate (Phys. Rev. X 5, 041041). Accelerating the path of discovery from theoretical physics equation to a working and useful prototype implementation from which later user applications can be developed is an important research challenge.

Question 2 Cyberinfrastructure Needed to Address the Research Challenge(s) (maximum ~1200 words): Describe any limitations or absence of existing cyberinfrastructure, and/or specific technical advancements in cyberinfrastructure (e.g. advanced computing, data infrastructure, software infrastructure, applications, networking, cybersecurity), that must be addressed to accomplish the identified research challenge(s).

Our most important and precious asset are the brains and time of our researchers: graduate students, postdoctoral research scientists, staff scientists, and faculty. Cyber infrastructure has one purpose: helping our researchers be productive.

In the field outlined above, work progresses from theory (paper, pencil, blackboard, mathematica) to prototype (python) to test implementations (C++, MPI) to user applications (XSEDE, NERSC, thousands to tens of thousands of CPUs...). In the course of a researcher's project, multiple of these ideas are tried out, discarded, reworked, and revised. Any impediment to this flow stalls a researcher and distracts him/her from progressing in their project. In the context of cyberinfrastructure, three components are extremely important:

1. The availability of interactive small-scale prototype hardware, helping a researcher to scale up his research from the theory prototype to a first implementation (10-100 cores, MPI, C++). The present state is that either researcher have such hardware in-house (in which paying for it / renewing it is challenging), or researchers have it on XSEDE, in which case the unpredictability of availability (queue times? will XSEDE grant be awarded, even after NSF grant was funded and researchers have arrived?) impacts the researcher's most important asset: his time.
Proposed cyberinfrastructure needed to address this challenge: A 'guaranteed availability' cluster/supercomputer on XSEDE on which a small amount of time can be awarded at the time a proposal is funded, which guarantees that at any time, X cores (with X between 10 and 500) will be available, interactively, for such a project.
Constructing such a prototyping computer would need a paradigm shift in our supercomputing centers: rather than having 'maximum usage' as an evaluation criterion, 'maximum availability' would be the target. The benefits to NSF would be huge: rather than sponsoring many small individual machines in researchers' home institutions, hardware could be combined, shared, maintained, and sustainable.

2. The availability to scale up computer codes quickly. This is a frequent problem faced by algorithm developers that is NOT faced by other 'computational users': at the time a proposal is submitted, the final code does not exist. Thus it is not possible to provide scaling curves for those codes at the time of submission. Consequently, large scale computer time is only available when it's too late: the code has been written, tested, the results produced on a smaller machine, and the publication is well underway. A change in the XSEDE allocation system that delegates the allocation of resources to the program managers (who have overall science outcome in mind), rather than the XSEDE technical staff (who have optimal usage of their systems in mind), would alleviate this problem.

3. Rapid prototyping software gets vastly accelerated by the availability of stable, tested, documented, and actively maintained domain libraries. Good examples are libint (quantum chemistry) or ALPS/ALPSCore (quantum physics). Delegating components of the code to these libraries not only saves a developer from implementing things twice, but also delegates maintenance and problems to the maintainers of these libraries. Continued NSF support of reusable software components is therefore absolutely crucial. Stable and rewarding career paths for people maintaining these libraries full time have to be developed, as they are not following the standard academic career paths.

Question 3 Other considerations (maximum ~1200 words, optional): Any other relevant aspects, such as organization, process, learning and workforce development, access, and sustainability, that need to be addressed; or any other issues that NSF should consider.

Every researcher has to manage the mismatch between available people and available grant funds. This is nothing new. A computational researcher also has to manage a mismatch between people, grants, and computer time, as the allocation of computer time is detached from the allocation of NSF funds. A coupling of NSF computational resources to the scientific grant review/award process should be considered and would be beneficial.

Consent Statement